

蒙古语言文字数字资源建设与共享工程
信息处理用蒙古文相关标准
MGC/01-06

信息技术 传统蒙古文固定短语语义分类体系及 其标记集

Information technology — classification system of traditional Mongolian fixed
phrases and the tag set

内蒙古自治区民族事务委员会

内蒙古大学

2016年11月

目 次

目 次	I
前 言	II
信息技术 传统蒙古文固定短语语义分类体系及其标记集	- 1 -
1 范围	- 1 -
2 术语和定义	- 1 -
3 标记集的内容	- 1 -
4 词类标记代码的制定原则	- 1 -
5 语义分类标记代码	- 2 -
6 标记集	- 2 -

前 言

本标准是依据《现代蒙古语》（1964年）等主流语法著作为基础，规定了信息处理用的蒙古文固定短语语义分类体系及其标记集。

本标准由内蒙古自治区民族事务委员会提出。

本标准起草单位：内蒙古师范大学。

本标准起草人：青格乐图、斯琴吐、包萨如拉、萨仁其木格、巴音孟和、红英、杨晓雪、马梦圆

信息技术 传统蒙古文固定短语语义分类体系及其标记集

1 范围

本标准规定了信息处理用蒙古文固定短语语义分类标记集,只包括蒙古文固定短语语义分类的标记符号。

2 术语和定义

下列术语和定义适用于本标准。

2.1

蒙古文信息处理 Mongolian information processing

用计算机对蒙古文固定短语语义分类等信息进行输入、存储、输出、搭配、提取等。

2.2

标记单位 tagging unit

蒙古文信息处理中使用的固定短语,如复合词、固定词、成语等。

2.3

词类 parts of speech

词的语法分类,根据词法形态、句法功能、语义特征划分出来的类。

2.4

语义分类

根据语义场理论和语义层次理论分类出来的类。

2.5

复合词 compound words

由两个实词(有时也包括辅助意义的词)组成,语法和词汇上都构成一个单位的固定短语。

2.6

固定词 fixed expressions

由一个实词和一个虚词组成,连用成型且未转义的固定短语。

2.7

成语 idiom

由三个或三个以上词组成,充当一个句子成分的固定短语。

2.8

标记 tag

对文本中标记单位的类别进行标注的代码。

3 标记集的内容

本标准包括固定短语词类以及复合词语义分类标记等的标记符号。

4 词类标记代码的制定原则

词类标记有两个级,用于英文字母的大小写。

4.1 《第一级标记》的命名方法

用大写英文字母标记第一级。如：

Y (来自HQRSIY_A的Y一个字母)：复合词

4.2 《第二级标记》的命名方法

用小写英文字母标记第二级。如：

Yn (来自_noun)：复合名词

Yt (来自_time)：复合时位词

Kv (来自_verb)：动词性成语

5 语义分类标记代码

5.1 语义分类标记代码的制定原则

语义分类层级最长的有11个级，标记代码用了英文字母大小写、阿拉伯数字和罗马数字。

5.2 语义分类标记的命名方法

为了考虑语义分类的类别多、层级多的现象，大的上几个层级用英文字母的大写来标记，中间两个层级用英文字母的小写来标记，最下层级用罗马数字标记，最下层级的上一层级为阿拉伯数字标记。命名时采用了英文命名为主，蒙古文命名为辅的办法。

如：BMPTIRff1 I 是B：实体复合名词——M：个体——P：人——T：人性——I：生物特性——R：亲属关系——F：家庭——f：长辈——1：父母——I：父亲。例子：HADAM ABV(岳父、老丈人)

6 标记集

以下是标记集内容。右上角的数字表示该项所包含的构形附加成分或连接元音数。“-”表示与词干分写的构形附加成分。“…”表示省略。为了简化分类，将“动态词尾”放在“连接元音”部分。

本标记代码分为词类标记代码、构形附加成分标记代码、其他标记代码。

6.1 词类标记集

词类标记代码见表1。

1 固定短语词类标记代码

序号	语文现象/单位	标记代码		词类名称	代码说明
		第一级	第二级		
1		Y		复合词	
2	AJV AHVI		Yn	复合名词	
3	GQYQ SAYIHAN		Ya	复合形容词	
4	ENE TERE		Yr	复合代词	
5	JQHIYAN BAYIGVL		Yv	复合动词	
6	DEGER_E DQQR_A		Yc	复合时位词	
7	HALTV MOLTO		Yd	复合副词	
8	GAR HATAGV	X		习用词	
9		K		成语	
10	HAGVCIN PVV-YIN HVGVRHAI		Kn	名词性成语	
11	LAI-YIN DQQGVR VSV GUYULGE		Kv	动词性成语	
12	ENG-UN	J		惯用词	
13	HELE SINJILEL	NT		名词术语	

6.2 固定短语语义分类标记集

见附表。